

Automated Decision-Making Systems (ADS) Workgroup Meeting

August 26, 2021

Agenda for August 26, 2021 Meeting

Agenda

- 2:30 Welcome and administrative updates – Katy Ruckle
- 2:35 Proviso Details – Katy Ruckle
- 2:50 Risks and Benefits of Automated Decision-Making Systems: A Preliminary Framework for Guiding Evaluation – David Luxton
- 3:20 Workgroup Discussion – Workgroup members
- 4:00 Answers to Open Tasks – Assigned work group members
- 4:15 Open Discussion
- 4:30 Adjourn

O
P
D
P

Welcome and Administrative Updates

O
P
D
P

Administrative Updates

- Teams Channel is established for Workgroup members
- Added workgroup members but unable to add those not in Washington's Enterprise Active Directory. Submitted a helpdesk ticket to add external members. This should be resolved by our next meeting.
- New workgroup member introduction



**Automated
Decision-making
Systems Workgroup**

O
P
D
P

Proviso Details

O
P
D
P

Definition of ADS in Proviso

(f) For purposes of this subsection, "automated decision system" or "system" means any algorithm, including one incorporating machine learning or other artificial intelligence techniques, that

- uses data-based analysis or calculations to make or support
 - government decisions,
 - judgments, or
 - conclusions
 - that cause a Washington resident to be treated differently than another Washington resident in the nature or amount of governmental interaction with that individual including, without limitation:
 - benefits,
 - protections,
 - required payments,
 - penalties,
 - regulations,
 - timing,
 - application, or
 - process requirements.

Proviso details

Are there changes needed regarding the development, procurement, and use of ADS by state agencies?

If yes, what types of changes?

Development?

Procurement?

Use?

How can ADS be reviewed before adoption?

How can ADS be reviewed while in operation?

Audited to ensure ADS is fair, transparent, and accountable?

How can state ensure ADS does not improperly advantage or disadvantage particular residents?

O
P
D
P

Workgroup must examine:

- (i) When state agency use of automated decision making systems should be prohibited;
- (ii) When state agency use of artificial intelligence-enabled profiling systems should be prohibited;
- (iii) Changes in the procurement of automated decision systems,
 - including when the procurement must receive prior approval by the office of chief information officer;
- (iv) How to
 - review,
 - identify, and
 - audit systems the system prior to procurement and after placed into service
 - does not discriminate against an individual,
 - or treat an individual less favorably than another,
 - in whole or in part
 - to ensure that art, on the basis of one or more factors enumerated in RCW 49.60.010;
- (v) How to provide public notice when an automated decision system is in use and how to appeal such decisions;
- (vi) How automated decision system data should be stored and whether such data should be shared outside the system; and
- (vii) Other issues determined that are necessary to govern state agency procurement and use ADS

Risks and Benefits of Automated Decision-Making Systems: A Preliminary Framework for Guiding Evaluation – David Luxton

O
P
D
P

Presenter – Dr. David Luxton

- Clinical psychologist and research scientist
- Authored academic articles in fields of artificial intelligence, ethics, and psychological health
- Associate Professor in Psychiatry and Behavior Sciences at the University of Washington

Risks and Benefits of Automated Decision-Making Systems: A Preliminary Framework for Guiding Evaluation

David D. Luxton, PhD., M.S.

O
P
D
P

About Bias

- Bias is “prejudice (known or unknown) in favor of or against one idea, thing, person, or group over another, usually in a way considered to be unfair.”
- Types of bias (relevant for design and use of Automated Decision Support (ADS) systems):
 - Algorithmic bias/prejudice (systematic and repeatable errors in a computer system that create unfair outcomes)
 - Negative legacy (bias already present in the data used to train algorithms)
 - Nonconscious/Implicit bias, Ethnocentrism bias, etc.
 - Underestimation (when there’s not enough data for the model to make confident conclusions for some segments of the population).

ADS System Benefits (Potentially)

- More efficient than humans (handling of more data and complex information, finding patterns humans can't see, scalable)
- More accurate and consistent than humans, resulting in optimal outcomes (e.g., in predicting behavior, determining needs of people), and uniformity (fairness)
- Less biased than humans (better than human judgement because humans are biased)

ADS System Risks

- Bias (e.g., based on race, gender, or any other status)
 - Built into the system as part of design
 - Bias that increases over time as process of machine learning (algorithmic bias), etc.
- “Black box” problem – we don’t know how the system works/makes decisions
- Appearance of arbitrary decision-making, lack of legitimacy, or “human” perspective
- Public does not trust the system, administrators (e.g., the government, hospitals, etc.)
- Harm. Systems decisions result in threat to safety and well being (e.g., public health, public safety), and civil rights/freedoms.

(Calo, 2017; Luxton, 2016; Luxton, 2014;
Luxton & Hudlicka, 2021)

Design and Review Solutions

- Goal for ADS systems – maximize benefits, minimize risks
- Consider that biases can be built into systems prior to and in design of systems, and some systems may increase in bias over time
- Use best practices to mitigate bias in design and use of ADS systems
- Use standardized criteria to review the design/intended use of ADS systems...

Recommended Standardized Review

- Applicability: Define scope of what systems should be reviewed (what is applicable ADS, what is not)
- Is this process of review fair across applications/populations, etc.?
- Guidelines/Checklist
 - Is there adequate evidence of system design source data methodology/size is representative and adequate to address bias?
 - Was testing (e.g., test, re-test, independence, etc.) adequate?
 - Is there plan/methodology to address change in validity/reliability over time (e.g., process for re-testing and monitoring performance)?
 - Is there a process for changing/updating if bias is discovered later?
 - Is the design and purpose of the system adequately transparent?
 - Are concerns of affected populations addressed/represented?
 - Are there better/more feasible alternatives to the system/tool?
 - Is system function appropriately documented (including how bias is addressed/limitations noted) (Important for transparency and trust)
 - Is there a process for understanding how the systems makes decisions (e.g., audit)?

References

- Calo, R. (2017). Artificial Intelligence Policy: A Primer and Roadmap. Available at: https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Calo.pdf
- Kamishima, T., Akaho, S. & Sakuma, J. (2011). Fairness-aware Learning through Regularization Approach. 2011 11th IEEE International Conference on Data Mining Workshops. Available at: https://www.kamishima.net/archive/2011-ws-icdm_padm.pdf
- Luxton, D. D. (2014). Recommendations for the ethical use and design of artificial intelligent care providers. Artificial Intelligence in Medicine. 62(1), 1-10. <https://doi.org/10.1016/j.artmed.2014.06.004>
- Luxton, D. D., Anderson, S. L., & Anderson M. (2015). Ethical Issues and Artificial Intelligence Technologies in Behavioral and Mental Health Care. In D. D. Luxton (Ed). Artificial Intelligence in Behavioral and Mental Health Care. San Diego: Elsevier Academic Press.
- Luxton, D. D. & Hudlicka, E. (2021). Intelligent Virtual Agents in Healthcare: Ethics and Application Considerations. In F. Jotterand, M. Ienca and M Liang (Eds)., Ethics of Artificial Intelligence in Brain & Mental Health. Springer Nature, Switzerland.

Calo, R. (2017). Artificial Intelligence Policy: A Primer and Roadmap. Available at:

https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Calo.pdf

Kamishima, T., Akaho, S. & Sakuma, J. (2011). Fairness-aware Learning through Regularization Approach. 2011 11th IEEE International Conference on Data Mining Workshops. Available at:

https://www.kamishima.net/archive/2011-ws-icdm_padm.pdf

Luxton, D. D. (2014). Recommendations for the ethical use and design of artificial intelligent care providers. Artificial Intelligence in Medicine. 62(1), 1-10.

<https://doi.org/10.1016/j.artmed.2014.06.004>

Luxton, D. D., Anderson, S. L., & Anderson M. (2015). Ethical Issues and Artificial Intelligence Technologies in Behavioral and Mental Health Care. In D. D. Luxton (Ed). Artificial Intelligence in Behavioral and Mental Health Care. San Diego: Elsevier Academic Press.

Luxton, D. D. & Hudlicka, E. (2021). Intelligent Virtual Agents in Healthcare: Ethics and Application Considerations. In F. Jotterand, M. Ienca and M Liang (Eds)., Ethics of Artificial Intelligence in Brain & Mental Health. Springer Nature, Switzerland.

Workgroup Discussion

O
P
D
P

ACLU Topic for Discussion

In the discussions leading to the ADS proviso, legislators identified one of the major impediments to legislative consideration of the issue as a simple lack of knowledge of ways in which Washington departments are using ADS. From the reports in that first meeting, it appears that pulling together high-level information about the systems was not too strenuous for the presenters (in most cases the departments had already looked at all the issues we were seeking information about). In view of that, we think it would be useful to the Work Group, but even more to the legislators if we could request that a range of other agencies share basic information (similar to that presented in our first meeting) about some of the ADS systems they use. An appendix to our report helping to put the issue in context with respect to the workings of Washington departments would be extremely useful.

ADS Budget Proviso Implementation

Questions for State Agencies

1. What is the automated decision system's name, vendor and version?
2. What decisions is the system utilized to make?
3. What data is inputted into the system?
4. How is the input data gathered, how often is it updated, and are subjective inputs ever audited for consistency across data collectors?
5. Is the decision algorithm is available for examination by the agency and/or the public?
6. Has there been any public or community engagement used in selection or design of the system? If so, please describe this engagement.
7. Does law or regulation mandate any of the decision system criteria? If so, which criteria?
8. Do the system's decisions intentionally differentially affect members of protected classes, such as selecting persons with disabilities for certain benefits?
9. Has the system has been tested for unintended bias by the agency or an independent third party? If so, what were the results? Describe briefly the nature of the testing.

ADS Budget Proviso Implementation

Questions for State Agencies cont...

10. Has the system has produced known erroneous results and if so, what were those errors (including the results of any audits conducted to check for erroneous results)?
11. In addition to any intentional differential effect on members of a protected class, are there are other differential effects on protected classes as shown by comparison of the system's data to general census data or, where relevant, subpopulation data, such as the effect on justice system defendants of color as contrasted with all defendants? If audits have been performed to determine such differential effects, what were the results of those audits?
12. Can those affected by a system decision review and challenge the basis for that decision? If so, how, and what were the results of any such challenges?
13. Is the decision system is operated by a third party? If so, what rules govern such operation and what audits are conducted to ensure compliance;
14. What is the fiscal impact of the system, including initial cost, operating costs, and any cost savings established as flowing from use of the system.
15. What were the personnel hours required to gather the relevant information (questions 1-14) for the system examined?

Federal Gov't Accountability Office (GAO) Artificial Intelligence Framework

- **Governance** - To help entities promote accountability and responsible use of AI systems, GAO identified key practices for establishing governance structures and processes to manage, operate, and oversee the implementation of these systems.
- **Data** - To help entities use data that are appropriate for the intended use of each AI system, GAO identified key practices to ensure data are of high quality, reliable, and representative.
- **Performance** - To help entities ensure AI systems produce results that are consistent with program objectives, GAO identified key practices for ensuring that systems meets their intended purposes.
- **Monitoring** - To help entities ensure reliability and relevance of AI systems over time, GAO identified key practices for monitoring performance and assessing sustainment and expanded use.
- <https://www.gao.gov/assets/gao-21-519sp.pdf>

Answers to Open Tasks – Assigned work group members

O
P
D
P

Action Item*	Description	Person Responsible	Resolution
01.06	Seek clarification on the level of access WSIPP may have to the WA ONE weights and algorithm in their evaluation of bias.	David Luxton	WSIPP has full access to the parameters (scores) they need to conduct the WA ONE analyses.
03.01	Share the joint Washington Secretary of State and CIO guidance related to data ownership best practices with workgroup members.	Elena McGrew	Received Guidance document from Elena. Will post with 8-26-21 Meeting Materials.
03.02	Provide the workgroup with reading recommendations, including toolkits from Berkley and the University of Washington.	Maria Angel	Resources posted in Teams Channel Wiki and also in 8-12-21 chat and slide deck.
03.03	Update Katy on his designation as a workgroup member representing the Commission on Hispanic Affairs or the ACLU of Washington.	Eric Gonzalez	9/9/21
03.04	Add identified topics for discussion to future workgroup agendas	Katy Ruckle	8/26/21
03.05	Send out prompts for what the workgroup may want to tackle for policy recommendations.	Katy Ruckle	8/26/21

Task 3.02 Resolution

- Danielle Keats Citron, *Technological Due Process*, 85 Wash. U. L. Rev. 1249 (2008).
https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview
- Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 Geo. L.J. 1147 (2017). https://scholarship.law.upenn.edu/faculty_scholarship/1734/
- Andrew Tutt, *An FDA for Algorithms*, 69 Admin. L. Rev. 83 (2017).
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994
- Emily Berman, *A government of Laws and not of Machines*, 98 B.U. L. Rev. 1277 (2018).
<https://www.bu.edu/bulawreview/files/2018/10/BERMAN.pdf>
- Brian Sheppard, *Warming up to inscrutability: How technology could challenge our concept of law*, 68 University of Toronto Law Journal 36 (2018). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2992497
- Mireille Hildebrandt, *Law as Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics*, 68 University of Toronto Law Journal 12 (2018).
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2983045
- Kenneth A. Bamberger & Deirdre K. Mulligan, *Procurement as Policy: Administrative Process for Machine Learning*, 34 Berkeley Technology Law Journal 781 (2019).
- Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 Emory Law Review 797 (2021). <https://scholarlycommons.law.emory.edu/elj/vol70/iss4/1/>
- Link to the 3 toolkits that have been created by the Tech Policy Lab to design tech policy: <https://techpolicylab.uw.edu/publications-resources/>

Articles from Jon Pincus:

- [If You Don't Trust A.I. Yet, You're Not Wrong](#), by Frank Pasquale in the New York Times, actually mentions Washington state's work! He also links to [California's similar bil, AB 13](#). His main point though is that we can learn a lot from the [proposed EU regulations](#) -- for example, the principle "that some uses of technology are simply too harmful to be permitted" should drive the global debate, and that some forms of AI are inherently high risk. Frank's very thoughtful about these issues (his 2015 book *The Black Box Society* is hugely important) so these are good perspectives to keep in mind.
- [Hundreds of AI tools have been built to catch covid. None of them helped.](#), by Will Douglass Haven in MIT Technology Review, is a sobering look at the hundreds of predictive tools that were developed. "None of them made a real difference, and some were potentially harmful." Still, some of them are being used in hospitals, and others are being marketed by developers. The article discusses some of the reasons why (and has links off to more detailed reports). This is useful both as a benchmark on the state of the art and for real-world examples of some of the kinds of things that need to be considered before systems are deployed, looking both at accuracy and at bias.
- <https://www.opengovpartnership.org/wp-content/uploads/2021/08/executive-summary-algorithmic-accountability.pdf> analysis (from The Ada Lovelace Institute (Ada), AI Now Institute (AI Now), and Open Government Partnership (OGP)), looking at regulation and policies to date in Western Europe and the US.

Open Discussion

O
P
D
P

Thank you!

O
P
D
P